

Multi-focus image fusion using maximum symmetric surround saliency detection

Durga prasad Bavirisetti* and Ravindra Dhuli

School of electronics engineering,

VIT University, Vellore, INDIA

bdps1989@gmail.com*, ravindrathuli@vit.ac.in

Received 4th Aug 2015; accepted 18th Dec 2015

Abstract

In digital photography, two or more objects of a scene cannot be focused at the same time. If we focus one object, we may lose information about other objects and vice versa. Multi-focus image fusion is a process of generating an all-in-focus image from several out-of-focus images. In this paper, we propose a new multi-focus image fusion method based on two-scale image decomposition and saliency detection using maximum symmetric surround. This method is very beneficial because the saliency map used in this method can highlight the saliency information present in the source images with well defined boundaries. A weight map construction method based on saliency information is developed in this paper. This weight map can identify the focus and defocus regions present in the image very well. So we implemented a new fusion algorithm based on weight map which integrate only focused region information into the fused image. Unlike multi-scale image fusion methods, in this method two-scale image decomposition is sufficient. So, it is computationally efficient. Proposed method is tested on several multi-focus image datasets and it is compared with traditional and recently proposed fusion methods using various fusion metrics. Results justify that our proposed method gives stable and promising performance when compared to that of the existing methods.

Key Words: Saliency map; Weight map; Out of focus; Multi-focus; Image fusion;

1. Introduction

Single image capture may not always convey complete information about the targeted scene. For better visual understanding of the scene two or more images have to be captured and all the useful information from these image captures should be combined into a single image. Image fusion [1] is the process of combining useful information from two or more images of the same scene into a single image. It has numerous applications in various fields such as digital photography [2], navigation [3], [4], military [4], [5], concealed weapon detection [3], [4], remote sensing [6] and medical imaging [4], [5] etc.

A generic classification [7] of image fusion is: single sensor image fusion (SSIF) and multi sensor image fusion (MSIF). In SSIF, single sensor is used to acquire multiple captures of the targeted scene whereas MSIF uses multiple sensors for the same purpose. These multiple images provide diverse information of the

Correspondence to: bdps1989@gmail.com

Recommended for acceptance by Janne Heikkilä

<http://dx.doi.org/10.5565/rev/elcvia.793>

ELCVIA ISSN: 1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

scene. It is very difficult to understand the scene from these multiple captures. So a single image should be generated out of these several images to provide more visual information than any one of the individual source images. Digital photography applications (Multi-focus fusion [8] and multi-exposure fusion [9]) come under SSIF category whereas applications such as concealed weapon detection, navigation, medical imaging, military, remote sensing fall under MSIF class.

Single sensor is not able to focus more than one object present in the scene at the same time because of inherent system limitations. So, several images with different focuses have to be captured and an all-in-one focus image has to be generated from these multiple images by the fusion process. Similarly, over and under exposed images provide visually less information so a well-exposed image can be generated by the process of fusion.

In MSIF applications, source images are captured by using different modalities. For example in medical imaging, modalities like computer tomography (CT), Magnetic resonance imaging (MRI) are used to capture complementary information of the targeted scene of a human body. CT provides hard tissue information (bone structure) whereas MRI gives the soft tissue information. The useful information from these complementary images should be integrated into a single image for better diagnosis and treatment.

Fusion process can be done at three levels [1] namely pixel, feature and decision. At pixel level, fusion is performed on each source image pixel by pixel. At feature level, features are derived from each image and fusion is performed on those extracted features. Decision level is a high level fusion. At this level, local decision makers are derived from each feature of an image and fusion is executed on probabilistic decision information derived from local decision makers. Pixel level fusion process is simple and very effective compared to other levels of fusion schemes. In this paper we concentrate only on pixel level fusion scheme.

An ideal pixel level fusion scheme exhibits the following properties [1]:

1. Fusion scheme must transfer all of the useful information from source imagery to the fused image.
2. It should not contain information loss of source imagery.
3. It should not introduce extra information or artifacts into the fused image.

In this context, for the past few decades several image fusion schemes have been implemented.

2. Literature

Pixel level fusion schemes are broadly classified as spatial and transform domain schemes. This classification is based on the domain in which the fusion algorithm is processing the source images. Now we briefly present spatial domain fusion schemes [7]. Pixel-wise manipulations such as average, maximum and minimum methods are the easiest methods in this class but these methods may produce brightness or colour distortions in the fused image. Principal component analysis (PCA) based fusion method [10] is computationally efficient but this method may not assure desirable results for most of the image datasets. Intensity-hue-saturation [11], brovey [12], bilateral gradient sharpness criterion (BGS) [13] methods are some of the successful methods in this category. But each of these methods have their own limitations.

Next category of interest is transform domain based fusion methods [14]. Pyramid based methods are most successful methods around 1990. Gradient (Grad), Laplacian, Filter-substrate-decimate (FSD), contrast, morphological difference pyramids are successfully used for the purpose of fusion [14]. However, these methods may produce artifacts around boundaries in the fused image.

Later it has been observed that discrete wavelet transform (DWT) [10] gives desirable decompositions compared to pyramids for the fusion purpose. But DWT is a shift variant method which may cause artifacts in the fused image. To solve this problem shift invariant discrete wavelet transform (SIDWT) [15] is used for fusion. Fusion process can also be implemented using recent transforms like curvelet [6], nonsubsampling contourlet [8], discrete cosine harmonic wavelet (DCHWT) [16], directional discrete cosine (DDCT) [17] transforms etc. All of the above mentioned fusion schemes come under a special category called as multi-scale decomposition (MSD) based fusion methods [2]. In MSD methods, each source image is decomposed into base layer containing large scale variations and successive detail layers containing small scale variations. Either manipulated base or manipulated detail layers of source images are combined to get the fused image. Anisotropic diffusion [9], [18], cross bilateral filter [3], weighted least square filter [19], guided

image filter [5], multi-resolution singular value decomposition (MSVD) [20], and higher order singular valued decomposition [21], fusion methods are the latest methods in this class. Most of these multi-scale fusion methods require multiple image decompositions to get satisfactory fusion results. Hence they are computationally expensive.

To solve above mentioned problems we propose a new image fusion method based on two-scale image decomposition and saliency map detection using maximum symmetric surround. The advantages of the algorithm are as follows:

- Unlike most of the MSD fusion methods proposed method uses two-scale image decomposition. For this purpose a simple average filter is employed. So it is computationally simple.
- In multi-focus images, focused region contains more salient information compared to defocused regions. So, saliency map should properly highlight the regions of interest from each source image. The saliency map used in this paper i.e, maximum symmetric surround saliency (MSSS) map can highlight the salient regions with well defined boundaries and few artifacts.
- A new weight map calculation based on saliency map is proposed. This weight map can highlight the complementary information from each source image.

The organization of the remaining paper is as follows: Section 3 reviews the saliency extraction algorithm. Section 4 explains the proposed methodology. Section 5 briefs various fusion metrics. Section 6 presents the experimental setup. Section 7 discusses the results. Section 8 concludes the paper.

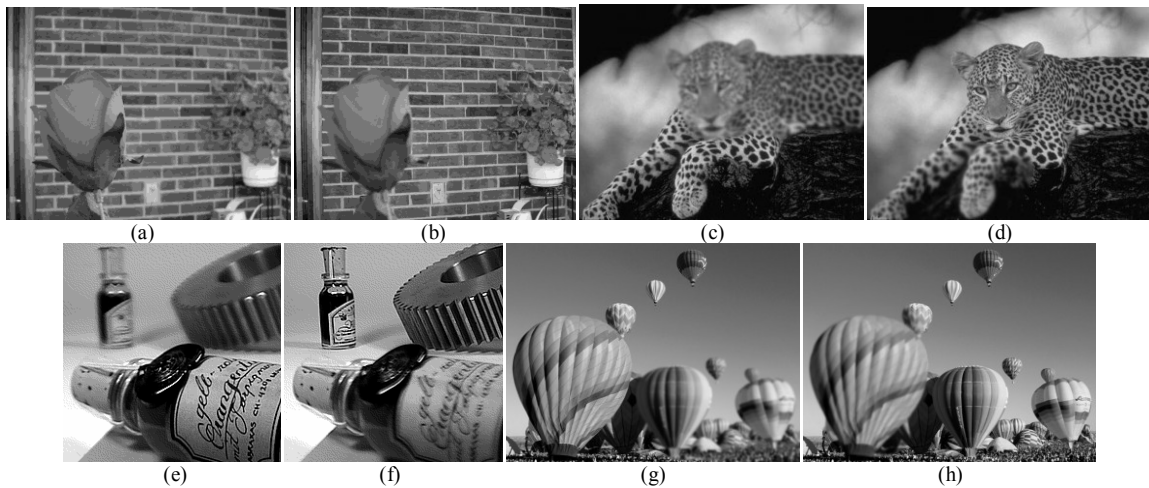


Fig.1: Multi-focus image data sets: (a), (b) are left, right focused images of a flower dataset. (c), (d) are foreground, background focused images of a leopard dataset. (e), (f) are foreground, background focused images of a bottle dataset and (g), (h) are left, right focused images of a parachute dataset.

3. Saliency map extraction

Saliency detection (SD) [22] is the process of detecting and highlighting visually significant regions which drag the human visual attention compared to other regions present in the image. SD is useful in many applications such as object segmentation, object reorganization and adaptive compression. However, in this paper we used SD for fusion purpose. A good SD method exhibits the properties mentioned below.

- It should highlight largest salient regions than smallest regions.
- It should uniformly highlight salient regions.
- Boundaries need to be well defined.
- It should ignore texture or noise artifacts.

In this view, so far many SD methods have been proposed.

SD algorithms [23-27] produce low resolution saliency maps. Some SD algorithms [24, 25, 27] generate ill-defined object boundaries. The saliency maps of these methods are not useful to generate weight maps for

the purpose of fusion because of these limitations. Achanta et. al [28] proposed a frequency tuned SD method which overcome the limitations of these existing saliency methods. This SD method is able to generate uniformly highlighted full resolution saliency maps with well-defined boundaries. However this method fails, if image consists of complex background or large salient regions. To solve above mentioned problems Achanta et al. proposed another SD algorithm called maximum symmetric surround SD method [22] which can highlight the salient object along with well-defined boundaries.

We prefer MSSS algorithm for fusion compared to other SD methods because

- 1) It gives saliency maps with full resolution and well defined boundaries.
- 2) Salient regions are calculated based on symmetric surrounds. Hence it can effectively highlight the salient regions in images with complex background.

In multi-focus images, focused regions provide more visual information than defocused regions because focused regions are more salient than defocused regions. So, these salient regions (visually significant regions) should be identified from the source images using SD algorithms. From the above discussion, we preferred SD using MSSS [22] for multi-focus image fusion.

The theory behind this SD algorithm is as follows:

At first, Achanta et al. proposed a frequency-tuned saliency detection algorithm [28] to utilize almost all low frequency content and most of the high frequency content to obtain perceptually good saliency maps with full resolution. This Saliency map is obtained by taking the Euclidean distance between the average of an image I_μ and each pixel of the Gaussian blurred version $I_f(u, v)$ of the same image.

$$S(u, v) = \|I_\mu - I_f(u, v)\|, \quad (1)$$

where $S(u, v)$ is the saliency map at a pixel location (u, v) . Gaussian blur of size 3×3 is chosen to get $I_f(u, v)$.

But this method fails when source image contains complex background. It highlights the background along with the salient object because this method treats entire image as the common surround for all pixels in the image. This is not desirable because to detect a pixel at the center of the salient object it should contain small lower cut-off frequency or it should contain high lower cut-off frequency to detect a pixel near boundary. So as we approach image boundaries we should use local surround regions instead of common surround regions to detect a pixel. This can be done by defining surround symmetry around the center pixel of its own sub image near the boundary. This process can increase the lower cut-off frequency. MSSS saliency map detection [22] of an image I of width w and height h , from eq(1) is defined as:

$$S_{ss}(u, v) = \|I_\mu(u, v) - I_f(u, v)\|, \quad (2)$$

where $I_\mu(u, v)$ is the average of the sub image at central pixel (u, v) and it is given as:

$$I_\mu(u, v) = \frac{1}{A} \sum_{i=u-u_0}^{u+u_0} \sum_{j=v-v_0}^{v+v_0} I(i, j), \quad (3)$$

where, u_0, v_0 represent off-sets and A indicates the area and is computed as:

$$\begin{aligned} u_0 &= \min(u, w-u), \\ v_0 &= \min(v, h-v), \\ A &= (2u_0 + 1)(2v_0 + 1). \end{aligned} \quad (4)$$

The sub images obtained using eq(3) and eq(4) are the maximum symmetric surround regions for a given central pixel. In multi-focus images, focused regions provide visually more information than defocused regions. In other way, focused regions are more salient than defocused regions. So we need to detect salient regions from these out-of-focus images using SD algorithms. We observe that the MSSS saliency detection algorithm is able to extract salient regions of the multi-focus images. The multi-focus datasets used for

simulations are shown in Fig. 1 and their corresponding saliency maps are displayed in Fig. 2. The process of saliency extraction using MSSS algorithm is denoted as:

$$S = MSSS(I), \quad (5)$$

where I is the input image and S is the output saliency map.

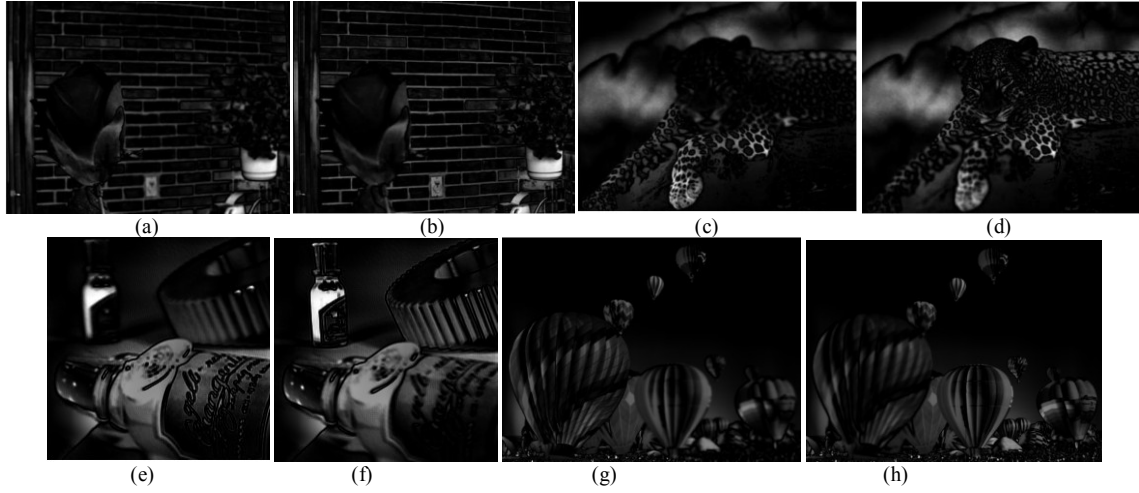


Fig. 2: Saliency maps of multi-focus image data sets: (a), (b) are saliency maps of a flower dataset . (c), (d) are saliency maps of a leopard dataset. (e), (f) are saliency maps of a bottle dataset and (g), (h) are saliency maps of a parachute dataset.

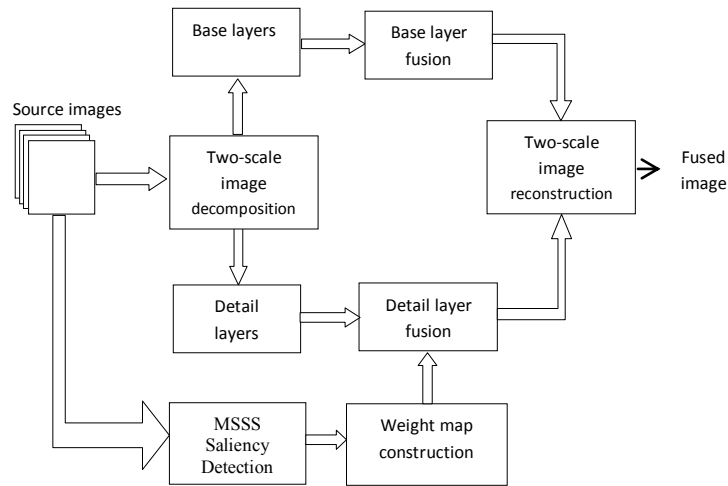


Fig. 3: Block diagram of the proposed method

4. Proposed method

The key idea of the proposed method is illustrated in the block diagram (Fig. 3). It is summarized in the following steps:

- A. Decompose source images into base and detail layers using an average filter.
- B. Calculate saliency maps of each source image using MSSS detection algorithm.
- C. Compute weight maps from extracted saliency map of each source image.
- D. Scale detail layers with these weight maps and combine all the scaled detail layers to obtain final detail layer.
- E. Compute final base layer by taking the average of all the base layers.
- F. Take linear combination of final base and detail layer to get the fused image.

This algorithm is explained in-detail in sub-sequent sub sections.

A. Two-scale image decomposition

Let us consider co-registered source images $\{I_n(x,y)\}_{n=1}^N$ of same size $p \times q$. These N -images are decomposed into base layer B containing large scale variations and detail layer D containing small scale variations as:

$$B_n = I_n * A, \quad (6)$$

where, A is an average filter of size w . The convolution operation indicated by $*$. The detail layers are obtained by subtracting base layers B_n from their corresponding source images I_n .

$$D_n = I_n - B_n. \quad (7)$$

here D_n is the detail layer.

B. Saliency detection algorithm

Saliency information from out-of-focus images is extracted using MSSS detection algorithm [22]. This algorithm is reviewed in section 3. The process of saliency extraction from source images I_n is represented as:

$$S_n = \text{MSSS}(I_n), \quad (8)$$

where, S_n is the saliency map of n^{th} source image.

C. Weight map calculation

In digital photography, each multi-focus image provides information about a particular focused region. We need to integrate all the focused regions into a single fused image. This can be done by properly choosing weight map of each source image. These weight maps should highlight the focused and defocused regions of the source images. Fig.4 shows the weight maps of various image datasets flower, leopard, bottle and parachute. These weight maps represent the complementary information i.e, focused and defocused regions. For example, as shown in Fig.4 weigh maps of focused and defocused regions are highlighted in red and green rectangles respectively represent the complementary information. These weight maps are calculated by normalizing the saliency maps as follows:

$$w_i = \frac{S_i}{\sum_{n=1}^N S_n}, \quad \forall i = 1, 2, \dots, N \quad (9)$$

D. Detail layers fusion

Here, detail layers are scaled with help of weight maps w_n calculated from MSSS detection algorithm and these scaled detail layers are combined to get the final detail layer \bar{D} as shown below:

$$\bar{D} = \sum_{n=1}^N w_n D_n. \quad (10)$$

E. Base layer fusion

Final base layer is generated by taking the average of base layers as:

$$\bar{B} = \frac{1}{N} \sum_{n=1}^N B_n. \quad (11)$$

F. Two-scale image reconstruction

Fused image is synthesized by taking the linear combination of \bar{B} and \bar{D} .

$$F = \bar{B} + \bar{D}. \quad (12)$$

5. Performance evaluation metrics

An image fusion algorithm can be evaluated from the visual quality and by using a set of fusion metrics. We consider traditional image fusion metrics such as average pixel intensity (*API*) [29], Mutual information (*MI*) [30], entropy (*H*) [31], fusion symmetry (*FS*) [1] and normalized correlation (*NC*) [29]. Their mathematical expressions are given below.

Let us consider a fused image $F(m, n)$ of size $p \times q$.

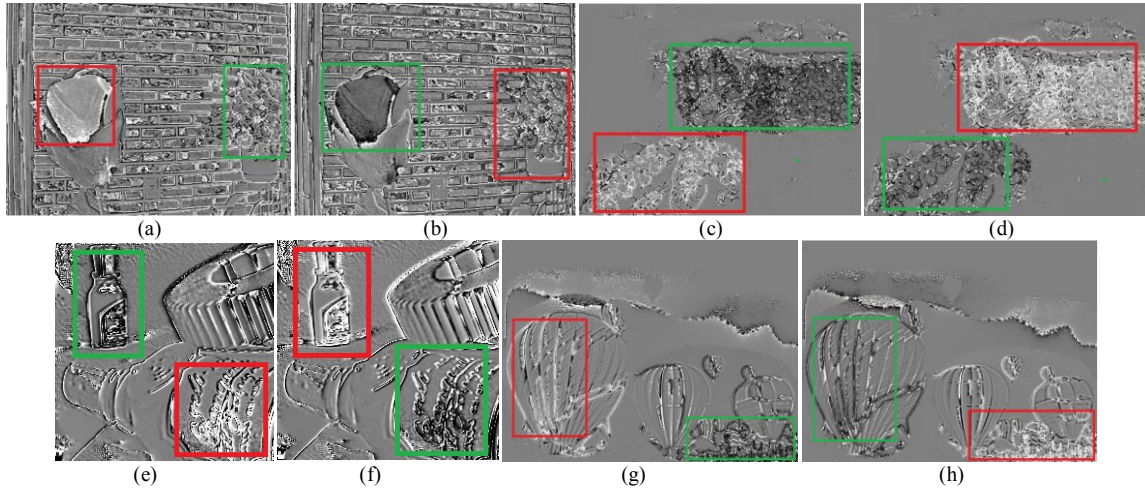


Fig. 4: Weight maps of various image data sets (red rectangle indicates the focused region, green rectangle indicates the defocused region): (a), (b) are weight maps of left, right focused images of a flower dataset. (c), (d) are weight maps of foreground, background focused images of a leopard dataset. (e), (f) are weight maps of foreground, background focused images of a bottle dataset and (g), (h) are weight maps of left, right focused images of a parachute dataset.

A. Average pixel intensity (*API*):

It measures the contrast of an image [29].

$$\mu = \frac{\sum_{m=1}^p \sum_{n=1}^q F(m, n)}{pq}, \quad (13)$$

B. Mutual information (*MI*):

It measures the overall information transferred from source images to the fused image [30] and it is given by:

$$MI = MI_{XF} + MI_{YF}, \quad (14)$$

where $MI_{XF} = \sum_m \sum_n p_{X,F}(m,n) \log_2 \left(\frac{p_{X,F}(m,n)}{p_X(m)p_F(n)} \right)$ is the mutual information between source image X and fused image F . Here $p_X(m)$ and $p_F(n)$ indicate the probability density functions of source images X and F respectively. $p_{X,F}(m,n)$ is the joint probability density function of source image X and the fused image F . Similarly MI_{YF} is the mutual information between Y and F .

C. Entropy (H):

Information quality index, entropy [31] is given by

$$H = - \sum_{l=0}^{L-1} p_l \log_2(p_l), \quad (15)$$

where p_l is the probability of intensity value l .

D. Fusion symmetry (FS):

It indicates the symmetric nature of the fused image with respect to the source images [1] and it is computed as:

$$FS = 2 - \left| \frac{MI_{XF}}{MI} - 0.5 \right|. \quad (16)$$

If both the sources images are equally symmetrical to the fused image then the value of FS is closer to 2 then the quality of the fused image will be good.

E. Normalized correlation (NC):

It indicates the amount by which the resultant image is relevant to the given input images [29] and is measured as

$$NC = \frac{(r_{XF} + r_{YF})}{2}, \quad (17)$$

where $r_{XF} = \frac{\sum_m \sum_n (x(m,n) - \bar{X})(f(m,n) - \bar{F})}{\sqrt{\left(\sum_m \sum_n (x(m,n) - \bar{X})^2 \right) \left(\sum_m \sum_n (f(m,n) - \bar{F})^2 \right)}}$, is normalized correlation between source

image X and fused image F . Similarly, r_{YF} represents the correlation coefficient between source image Y and fused image F .

With these fusion metrics in hand, we evaluate the performance of the proposed algorithm. For better quality of the fused image all these fusion metrics have to be high.

6. Experimental setup

In this section, we discuss about the image database, various image fusion algorithms used for comparison and the analysis of free parameter (average filter window size w) for optimal performance of the proposed method.

A. Image database

Experiments are carried on various multi-focus image datasets however due to space constraint results are presented for 4 image datasets namely flower, leopard, bottle and parachute. These datasets are displayed in Fig. 1 for reference. These datasets are available at <http://home.ustc.edu.cn/~liuyu1/>.

B. Methods for comparison

Proposed method is compared with spatial domain fusion algorithms such as PCA [10], BGS [13] and transform domain fusion algorithms like Grad [14], FSD [14], SIDWT [14], DCHWT [16], DDCT [17] and MSVD [20]. Default settings are adopted for all of these methods.

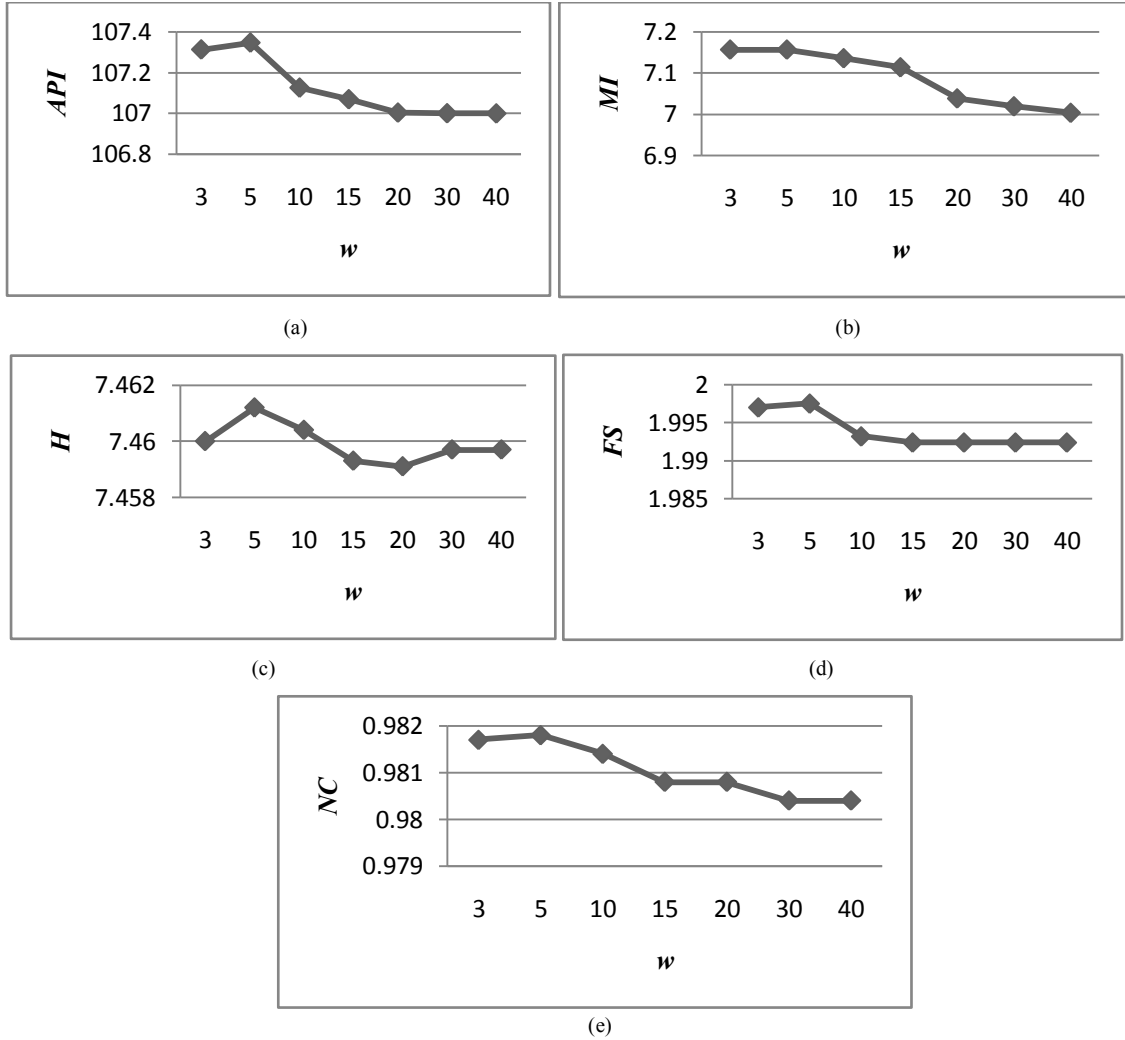


Fig. 5: Effect of W on fusion metrics (a) API , (b) MI , (c) H , (d) FS , (e) NC .

C. Free parameter analysis

The proposed method uses average filter to extract base and detail layer information from source images. Size of the average filter affects the performance of the proposed algorithm. So we have to tune the size of the average filter for the best performance. Fusion metrics API , MI , H , FS , NC are plotted in Fig. 5 by averaging over 4 multi-focus datasets. It can be observed that $w=5$ is the best choice.

7. Results and analysis

In this section a comparative analysis of the proposed method with different image fusion methods is done in terms of visual quality (qualitative analysis) and fusion metrics (quantitative analysis).

TABLE 1 : Quantitative analysis of different fusion methods of flower dataset

Metric	Method								
	<i>PCA</i>	<i>Grad</i>	<i>FSD</i>	<i>MSVD</i>	<i>SIDWT</i>	<i>BGS</i>	<i>DCHWT</i>	<i>DDCT</i>	<i>Proposed</i>
<i>API</i>	105.4159 (5)	105.1825 (8)	105.2029 (7)	105.5793 (4)	105.5807 (3)	105.4058 (6)	104.6864 (9)	106.5832 (1)	105.8115 (2)
<i>MI</i>	5.3496 (4)	4.6830 (8)	4.6400 (9)	5.0789 (7)	5.3512 (3)	5.3469 (5)	5.5972 (1)	5.2801 (6)	5.4405 (2)
<i>H</i>	7.1178 (7)	6.8987 (9)	6.8995 (8)	7.1307 (6)	7.1321 (5)	7.1325 (4)	7.1483 (2)	7.1660 (1)	7.1364 (3)
<i>FS</i>	1.9784 (2)	1.8981 (7)	1.8995 (6)	1.9569 (4)	1.9481 (5)	1.7860 (9)	1.8762 (8)	1.9618 (3)	1.9943 (1)
<i>NC</i>	0.9704 (5)	0.9627 (7)	0.9620 (8)	0.9735 (4)	0.9769 (2)	0.9595 (9)	0.9691 (6)	0.9737 (3)	0.9787 (1)

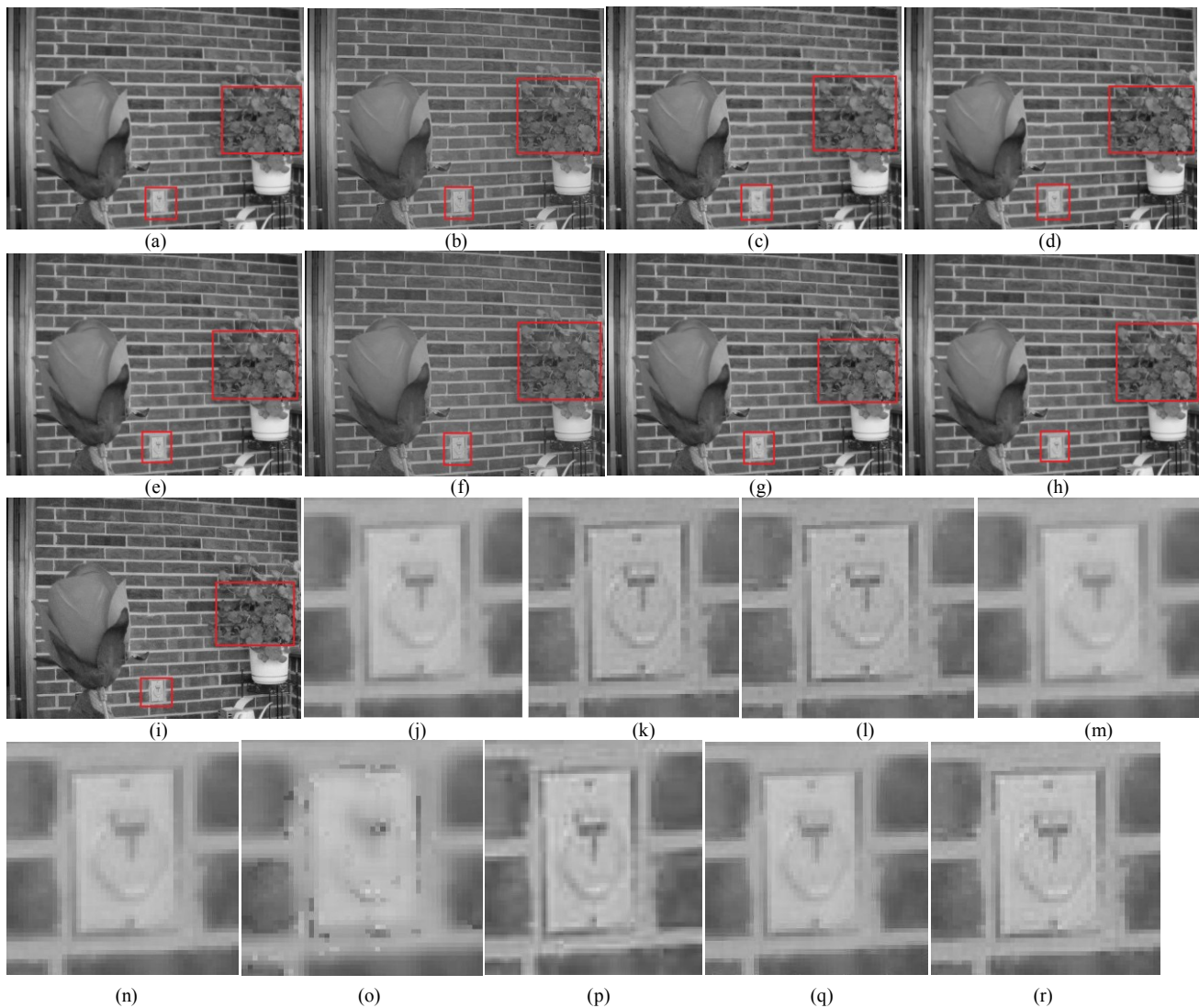


Fig. 6: Comparison of visual quality of fused images of various methods for flower dataset (a) PCA, (b) Grad, (c) FSD, (d) MSVD, (e) SIDWT, (f) BGS, (g) DCHWT, (h) DDCT, (i) Proposed. figures (j)-(r) show the zoom version of switch portion of (a)-(i) respectively.

TABLE 2 : Quantitative analysis of different fusion methods of leopard dataset

Metric	Method								
	PCA	Grad	FSD	MSVD	SIDWT	BGS	DCHWT	DDCT	Proposed
<i>API</i>	92.0730 (3)	92.0539 (8)	92.0668 (6)	92.0719 (4)	92.0699 (5)	91.1796 (9)	92.1767 (1)	92.0662 (7)	92.0736 (2)
<i>MI</i>	8.7521 (3)	5.7340 (8)	5.7262 (9)	6.5946 (7)	8.7109 (4)	8.0315 (6)	8.2713 (5)	8.7736 (1)	8.7523 (2)
<i>H</i>	7.4072 (8)	7.4349 (4)	7.4354 (3)	7.4067 (9)	7.4166 (7)	7.4208 (5)	7.4359 (2)	7.4203 (6)	7.4400 (1)
<i>FS</i>	1.9873 (4)	1.9840 (7)	1.9841 (6)	1.9975 (2)	1.9944 (3)	1.8790 (9)	1.9787 (8)	1.9863 (5)	1.9979 (1)
<i>NC</i>	0.9946 (2)	0.9891 (7)	0.9890 (8)	0.9908 (5)	0.9941 (3)	0.9893 (6)	0.9872 (9)	0.9925 (4)	0.9948 (1)

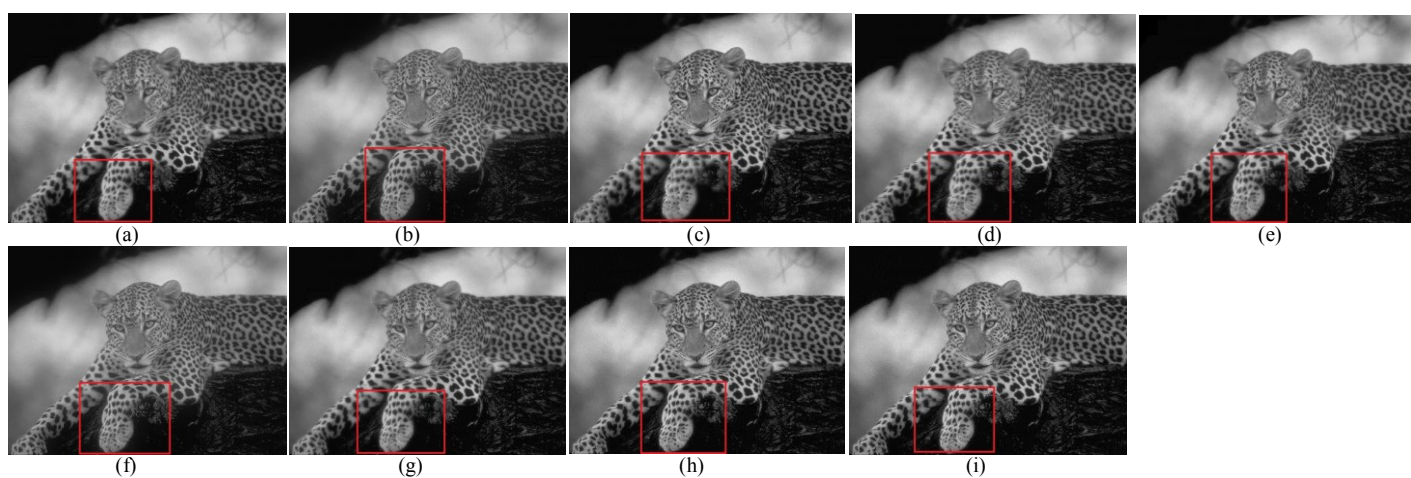


Fig. 7: Comparison of visual quality of fused images of various methods for leopard dataset (a) PCA, (b) Grad, (c) BGS, (d) MSVD, (e) SIDWT, (f) FSD, (g) DCHWT, (h) DDCT, (i) Proposed.

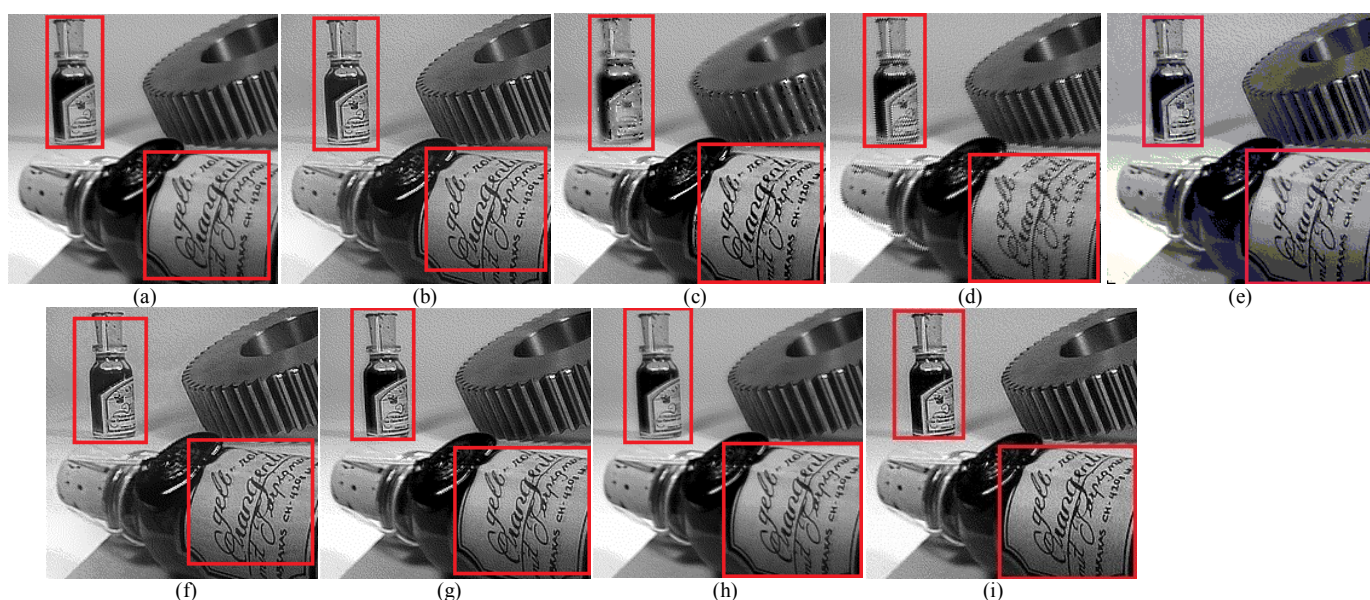


Fig. 8: Comparison of visual quality of fused images of various methods for bottle dataset (a) PCA, (b) Grad, (c) BGS, (d) MSVD, (e) SIDWT, (f) FSD, (g) DCHWT, (h) DDCT, (i) Proposed.

TABLE 3 : Quantitative analysis of different fusion methods of bottle dataset

Metric	Method								
	<i>PCA</i>	<i>Grad</i>	<i>FSD</i>	<i>MSVD</i>	<i>SIDWT</i>	<i>BGS</i>	<i>DCHWT</i>	<i>DDCT</i>	<i>Proposed</i>
<i>API</i>	117.6251 (6)	117.9222 (3)	117.9944 (1)	117.6989 (4)	117.6597 (5)	117.1480 (9)	117.5362 (7)	117.2105 (8)	117.9933 (2)
<i>MI</i>	4.2481 (6)	4.0358 (7)	4.0134 (8)	4.0129 (9)	4.3561 (5)	4.7148 (2)	4.6363 (3)	4.5529 (4)	4.9429 (1)
<i>H</i>	5.4125 (8)	7.7195 (2)	7.6604 (4)	7.3429 (7)	7.6651 (3)	3.6612 (9)	7.6048 (5)	7.5398 (6)	7.8119 (1)
<i>FS</i>	1.9979 (6)	1.9984 (5)	1.9991 (3)	1.9994 (2)	1.9988 (4)	1.8288 (8)	1.9946 (7)	1.9979 (6)	1.9996 (1)
<i>NC</i>	0.9606 (4)	0.9495 (6)	0.9472 (8)	0.9488 (7)	0.9618 (3)	0.9424 (9)	0.9679 (1)	0.9587 (5)	0.9632 (2)

TABLE 4 : Quantitative analysis of different fusion methods of parachute dataset

Metric	Method								
	<i>PCA</i>	<i>Grad</i>	<i>FSD</i>	<i>MSVD</i>	<i>SIDWT</i>	<i>BGS</i>	<i>DCHWT</i>	<i>DDCT</i>	<i>Proposed</i>
<i>API</i>	113.4778 (8)	113.4998 (5)	113.5008 (4)	113.4811 (6)	113.4781 (7)	113.5030 (3)	113.6250 (1)	113.4676 (9)	113.5050 (2)
<i>MI</i>	9.4903 (4)	7.2103 (8)	7.2051 (9)	9.0965 (7)	9.4294 (5)	9.2927 (6)	9.5255 (2)	9.5276 (1)	9.4906 (3)
<i>H</i>	7.4501 (6)	7.3468 (8)	7.3460 (9)	7.4527 (5)	7.4565 (4)	7.4427 (7)	7.4805 (1)	7.4695 (2)	7.4572 (3)
<i>FS</i>	1.9860 (8)	1.9918 (6)	1.9917 (7)	1.9981 (3)	1.9980 (4)	1.9038 (9)	1.9971 (5)	1.9988 (1)	1.9982 (2)
<i>NC</i>	0.9905 (3)	0.9803 (8)	0.9802 (9)	0.9885 (5)	0.9900 (4)	0.9835 (7)	0.9917 (1)	0.9877 (6)	0.9906 (2)

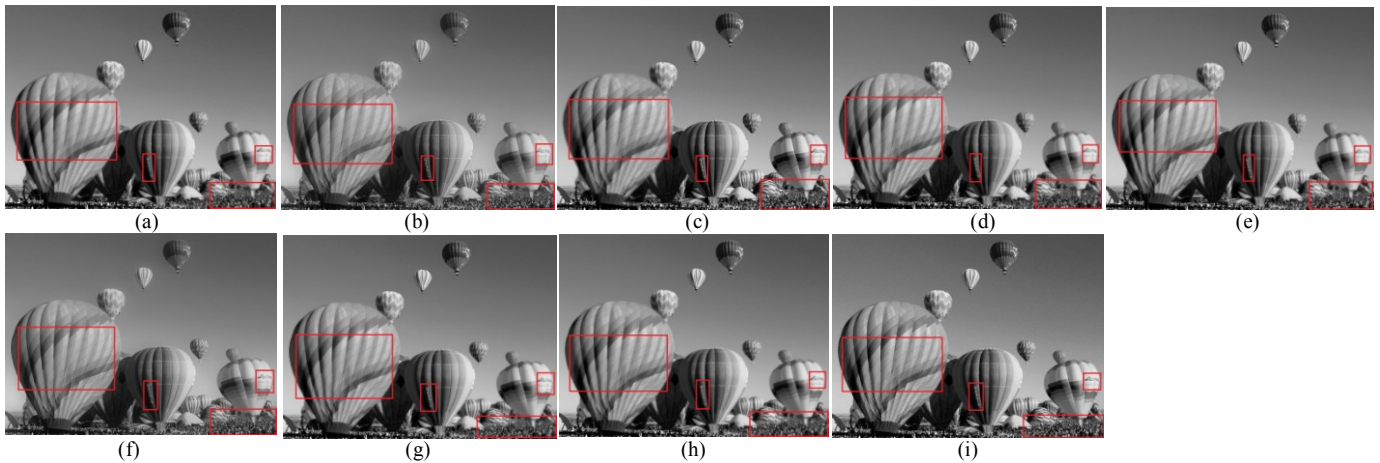


Fig. 9: Comparison of visual quality of fused images of various methods for parachutedataset (a) PCA, (b) Grad, (c) BGS, (d) MSVD, (e) SIDWT, (f) FSD, (g) DCHWT, (h) DDCT, (i) Proposed.

TABLE 5 : Average *API*, *MI*, *H*, *FS*, *NC* fusion metric values over four multi-focus image datasets of various fusion methods.

Metric	<i>PCA</i>	<i>Grad</i>	<i>FSD</i>	<i>MSVD</i>	Method <i>SIDWT</i>	<i>BGS</i>	<i>DCHWT</i>	<i>DDCT</i>	<i>Proposed</i>
<i>API</i>	107.148 (7)	107.1646 (6)	107.1912 (5)	107.2078 (3)	107.1971 (4)	106.8091 (9)	107.0061 (8)	107.3319 (2)	107.3459 (1)
<i>MI</i>	6.9601 (5)	5.4158 (8)	5.3962 (9)	6.1957 (7)	6.9619 (4)	6.8465 (6)	7.0076 (3)	7.034 (2)	7.1566 (1)
<i>H</i>	6.8469 (8)	7.350 (5)	7.3353 (6)	7.3333 (7)	7.4176 (2)	6.4143 (9)	7.4174 (3)	7.3989 (4)	7.4614 (1)
<i>FS</i>	1.9874 (3)	1.9681 (7)	1.9686 (6)	1.9880 (2)	1.9848 (5)	1.8494 (9)	1.9617 (8)	1.9862 (4)	1.9975 (1)
<i>NC</i>	0.9791 (3)	0.9704 (7)	0.9696 (8)	0.9754 (6)	0.9807 (2)	0.9687 (9)	0.9790 (4)	0.9782 (5)	0.9819 (1)

A. Qualitative analysis

Source images of 4 multi-focus image datasets flower, leopard, bottle and parachute are shown in Fig. 1. Their corresponding qualitative analysis is displayed in Figs. 6-9. In these figures, subfigures (a)-(i) illustrate fused images of PCA, Grad, FSD, MSVD, SIDWT, BGS, DCHWT, DDCT and the proposed method respectively.

Fig. 1(a) and 1(b) are left and right focused images of the flower dataset respectively. Left focused image provides the information of the focused flower region. Right focused image gives the information of sharpened switch and flower wage. By the process of fusion we can integrate focused information of these images into a single image. Figs. 6(a)-(i) are the fused images of PCA, Grad, FSD, MSVD, SIDWT, BGS, DCHWT, DDCT and proposed method respectively. As highlighted in Fig. 6(i), our proposed method has generated more focused regions (such as switch, flower wage and rose regions) compared to other methods. For an example, zoomed switch portion of fused images of different methods along with the proposed method are presented in Figs. 6 (j)-(r) for better visual understanding. BGS switch region (Fig. 6(o)) is completely blurred. Switch portions of PCA (Fig. 6(j)), MSVD (Fig. 6(m)), SIDWT (Fig. 6(n)) and DDCT (Fig. 6(q)) are partially blurred. Zoomed switch regions of Grad (Fig. 6(k)), FSD (Fig. 6(l)) are looking good. However, they are not providing enough details of the switch region. DCHWT switch region (Fig. 6(p)) is also looking better. But, this method introduces extra information (artifacts) into the switch region. Edges of the zoomed portion are distorted. It can be observed from the switch region (Fig. 6(r)) that the proposed method is able to integrate more focused regions with few artifacts compared to the state-of-the-art fusion methods.

Figs. 1(c) and 1(d) are foreground and background focused images of the leopard dataset respectively. Foreground focused image conveys the focused regions such as legs, claw of the leopard and a portion of the tree. Background focused image conveys the information of head, body regions of the leopard and the background. To know the entire information in the scene, all the focused regions have to be combined in the fused image. Figs. 7(a)-(i) give the visual display of various methods with the claw of the leopard highlighted in red rectangle. BGS fused image (for example claw region) (Fig. 7(c)) is visually distorted. Grad, FSD fused image results (Figs. 7(b) and 7(f)) are not satisfactory. Here DDCT fused image (Fig. 7(h)) is good. As shown in Figs. 7(a), 7(d), 7(e) and 7(g), PCA, MSVD, SIDWT and DCHWT fused images are partially blurred. As shown in Fig. 7(i) our proposed method is able to generate visually a good focused image for this dataset as well.

Foreground and background focused images of the bottle dataset are displayed in Figs. 1(e) and 1(f). Foreground focused image displays the focused bottle region and text on it. Background focused image displays the small bottle and a wheel. MFF can combine this diverse information into a single image. Fused images of various MFF methods with some regions of interest are highlighted with red rectangles in Figs. 8 (a)-(i). Foreground region of BGS fused image (Fig. 8 (c)) is well focused. But, it fails to incorporate well focused background content. For example, small bottle and wheel regions are completely distorted. MSVD fused image (Fig. 8(d)) is visually not good. SIDWT (Fig. 8(e)) introduces artifacts into the fused image. As shown in Figs. 8(b) and 8(f), Grad and FSD fused images are looking good but quality of these

images is not of satisfactory level. PCA, DCHWT and DDCT fused images (as displayed in Figs. 8(a), 8(g) and 8(h)) are able to integrate source information into the fused image. However, as shown in Fig. 8(i) the proposed method integrates bottle in foreground and small bottle, wheel in the background into the fused image with more visual clarity compared to that of the existing methods (Figs. 8(a)-(h)).

Left and right focused images of a parachute dataset are illustrated in Figs. 1(g) and 1(h) respectively. Left focused image pay attention to a big parachute on the left hand side. Right focused image shows the remaining parachutes on the right hand side of the image. Fused images of various methods are displayed in Fig. 9 and some regions such as a portion of big parachute, text on another two small parachutes and a part in the right bottom are highlighted in red rectangles in all of these fused images for better analysis. BGS fused image (Fig. 9(c)) contains only right side focused regions. It failed to integrate left side focused big parachute into the fused image. Grad and FSD fused images (Figs. 9(b) and 9(f)) contain both left and right focused regions. But they are not able to give sufficient information. As shown in Figs. 9(a), 9(d-e), 9(g-h)), fused images of PCA, MSVD, SIDWT, DCHWT and DDCT are looking good. However, as displayed in Fig. 9(i) proposed fused image gives more focused regions. For example, as shown in red rectangles on the right hand side of the image, text on the parachutes and right bottom regions are more visible and clear than remaining fused images.

B. Quantitative analysis

Fused image cannot be judged exclusively by seeing the fused image. So far we have discussed qualitative analysis. Now we will discuss the quantitative analysis of different fusion methods.

Comparison of fusion metrics API , MI , H , FS and NC for different methods along with the proposed method is presented in Tables. 1-4. Average values of API , MI , H , FS and NC over 4 image datasets are presented in Table. 5. Fusion metrics with highest value are highlighted in bold letters.

A single image fusion algorithm may not always give better performance in all fusion metrics for all the image datasets. An algorithm which gives better performance for one dataset may not give the same performance for another dataset. Some fusion algorithms may give better results for few fusion metrics alone. In other fusion metrics these methods may fail. So, fusion algorithm should be assessed by taking the performance of all the fusion metrics into the consideration.

It can be noted that in Tables. 1-4, no individual fusion algorithm is giving better performance in all fusion metrics for all image datasets. In addition, we can observe that no individual fusion algorithm has given the best fusion metric value for all the datasets. Performance (fusion metric values) of a fusion algorithm is changing from one dataset to other. So it is difficult to judge from the individual dataset evaluation.

We are considering two types of evaluation techniques for better assessment.

1) Ranking. 2) Average metric values.

Ranking:

We have ranked each fusion algorithm based on its performance. Fusion algorithms with highest fusion metric value is given rank 1. There are total nine fusion algorithms including the proposed method. So, nine ranks are given as shown in Tables. 1-4. It is easy to observe that the rank of our algorithm is good and stable with not much fluctuation, unlike other algorithms.

Average metric values:

Motivated by the analysis in recent papers [5], [32], [33] we have also considered the average value of the fusion metrics (calculated over 4 image datasets) in Table. 5. It obvious from this table, that our algorithm is giving superior performance in all fusion metrics compared to that of state-of-the-art fusion methods.

From these two criteria we can justify that our algorithm is promising and reliable independent of the choice of metric, dataset.

8. Conclusion

In this paper, a new multi-focus image fusion method based on two-scale image decomposition and visual saliency detection is proposed.

- Unlike most of the MSD fusion techniques, proposed method uses two-scale image decomposition to extract base and detail layers. So, it is computationally simple.
- A feature extraction process using MSSS Saliency detection is explored for multi-focus images. This SD method is able to detect the visually significant information of source images.
- A new weight map construction process based on visual saliency is proposed. The weight maps of source images are able to identify focus and defocus regions of the source images very well.
- Proposed method is compared both qualitatively and quantitatively with different fusion methods. Results justify that our method is giving superior performance than the existing methods.

References

- [1] Goshtasby, A. Ardesbir, and Stavri Nikolov. "Image fusion: advances in the state of the art." *Information Fusion* 8.2: 114-118, 2007.DOI:10.1016/j.inffus.2006.04.001
- [2] Zhang, Zhong, and Rick S. Blum. "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application." *Proceedings of the IEEE* 87.8 : 1315-1326, 1999.DOI:10.1109/5.775414
- [3] Kumar, BK Shreyamsha. "Image fusion based on pixel significance using cross bilateral filter." *Signal, Image and Video Processing* :1-12, 2013. DOI 10.1007/s11760-013-0556-9
- [4] Bavirisetti, Durga Prasad, and Ravindra Dhuli. "Multi Sensor Image Fusion Using Saliency Map Detection." *International Review on Computers and Software (IRECOS)* 10.7 : 757-763, 2015. DOI: 10.15866/irecos.v10i7.6793
- [5] Li, Shutao, Xudong Kang, and Jianwen Hu. "Image fusion with guided filtering." *Image Processing, IEEE Transactions on* 22.7 : 2864-2875, 2013.DOI: 10.1109/TIP.2013.2244222
- [6] Choi, Myungjin, et al. "Fusion of multispectral and panchromatic satellite images using the curvelet transform." *Geoscience and remote sensing letters, IEEE* 2.2 : 136-140, 2005. DOI: 10.1109/LGRS.2005.845313
- [7] Stathaki, Tania. Image fusion: algorithms and applications. Academic Press, 2011.
- [8] Zhang, Qiang, and Bao-long Guo. "Multifocus image fusion using the nonsubsampling contourlet transform." *Signal Processing* 89.7 : 1334-1346, 2009. DOI: 10.1016/j.sigpro.2009.01.012
- [9] Singh, Harbinder, Vinay Kumar, and Sunil Bhooshan. "Anisotropic diffusion for details enhancement in multiexposure image fusion." *ISRN Signal Processing*, 2013.DOI: 10.1155/2013/928971
- [10] Naidu, V. P. S., and J. R. Rao. "Pixel-level image fusion using wavelets and principal component analysis." *Defence Science Journal* 58.3 : 338-352, 2008. DOI: 10.14429/dsj.58.1653
- [11] Chavez, Pats, Stuart C. Sides, and Jeffrey A. Anderson. "Comparison of three different methods to merge multiresolution and multispectral data- Landsat TM and SPOT panchromatic." *Photogrammetric Engineering and remote sensing* 57.3: 295-303, 1991.
- [12] Pohl, Cle, and John L. Van Genderen. "Review article multisensor image fusion in remote sensing: concepts, methods and applications." *International journal of remote sensing* 19.5 : 823-854, 1998. DOI:10.1080/014311698215748
- [13] Tian, Jing, et al. "Multi-focus image fusion using a bilateral gradient-based sharpness criterion." *Optics communications* 284.1 : 80-87, 2011.DOI: 10.1016/j.optcom.2010.08.085
- [14] Rockinger, O. "Multiresolution-Verfahren zur Fusion dynamischer Bildfolgen", Ph.D. Thesis, Technische Universität Berlin, 1999, ISBN 3-933342-77-5.
- [15] Rockinger, Oliver. "Image sequence fusion using a shift-invariant wavelet transform." *Image Processing, 1997. Proceedings., International Conference on*. Vol. 3. IEEE, 1997.DOI:10.1109/ICIP.1997.632093
- [16] Kumar, BK Shreyamsha. "Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform." *Signal, Image and Video Processing* 7.6 : 1125-1143, 2013.DOI: 10.1007/s11760-012-0361-x

- [17] Naidu, V. P. S. "Hybrid DDCT-PCA based multi sensor image fusion." *Journal of Optics* 43.1 : 48-61, 2014. DOI: 10.1007/s12596-013-0148-7
- [18] Bavirisetti, D.P.; Dhuli, R., "Fusion of Infrared and Visible Sensor Images Based on Anisotropic Diffusion and Karhunen-Loeve Transform," *IEEE Sensors Journal* , 16.1 : 203-209, 2016. DOI:10.1109/JSEN.2015.2478655
- [19] Zhao, Jufeng, et al. "Fusion of visible and infrared images using saliency analysis and detail preserving based image decomposition." *Infrared physics & technology* 56 : 93-99, 2013. DOI: 10.1016/j.infrared.2012.11.003
- [20] Naidu, V. P. S. "Image fusion technique using multi-resolution singular value decomposition." *Defence Science Journal* 61.5 : 479-484, 2011. DOI: 10.14429/dsj.61.705
- [21] Liang, Junli, et al. "Image fusion using higher order singular value decomposition." *Image Processing, IEEE Transactions on* 21.5 : 2898-2909, 2012. DOI: 10.1109/TIP.2012.2183140
- [22] Achanta, Radhakrishna, and Sabine Süsstrunk. "Saliency detection using maximum symmetric surround." *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010. DOI: 10.1109/ICIP.2010.5652636
- [23] Ma, Yu-Fei, and Hong-Jiang Zhang. "Contrast-based image attention analysis by using fuzzy growing." *Proceedings of the eleventh ACM international conference on Multimedia*. ACM, 2003. DOI: 10.1145/957013.957094
- [24] Itti, Laurent, Christof Koch, and Ernst Niebur. "A model of saliency-based visual attention for rapid scene analysis." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 11 : 1254-1259, 1998. DOI:10.1109/34.730558
- [25] Frintrop, Simone, Maria Klodt, and Erich Rome. "A real-time visual attention system using integral images." *International conference on computer vision systems*, 2007.
- [26] Hou, Xiaodi, and Liqing Zhang. "Saliency detection: A spectral residual approach." *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007. DOI: 10.1109/CVPR.2007.383267
- [27] Harel, Jonathan, Christof Koch, and Pietro Perona. "Graph-based visual saliency." *Advances in neural information processing systems*, 2006.
- [28] Achanta, Radhakrishna, et al. "Frequency-tuned salient region detection." *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*. IEEE, 2009. DOI: 10.1109/CVPR.2009.5206596
- [29] Blum, Rick S., and Zheng Liu, eds. *Multi-sensor image fusion and its applications*. CRC press, 2005.
- [30] Cover, Thomas M., and Joy A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.
- [31] Leung, Lau Wai, Bruce King, and Vijay Vohora. "Comparision of image data fusion techniques using entropy and INI." *Paper presented at the 22nd Asian Conference on Remote Sensing*. Vol. 5, 2001.
- [32] Bai, Xiangzhi. "Infrared and visual image fusion through feature extraction by morphological sequential toggle operator." *Infrared Physics & Technology* 71: 77-86, 2015. DOI: 10.1016/j.infrared.2015.03.001
- [33] Cui, Guangmang, et al. "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition." *Optics Communications* 341 :199-209, 2015. DOI: 10.1016/j.optcom.2014.12.032